

Purposeful Language Assessment: Selecting the Right Alternative Test

By John M. Norris

Language teachers are often faced with the responsibility of selecting or developing language tests for their classrooms and programs. However, deciding which testing alternatives are the most appropriate for a particular language education context can be daunting, especially given the increasing variety of instruments, procedures, and practices available for language testing. Such alternatives include not only test types with long traditions of use—such as multiple choice, matching, true-false, and fill-in-the-blank tests; cloze and dictation procedures; essay exams; and oral interviews—but also tests differing in scope and structure from these well-known options. For example, technological developments have led to a number of new language testing formats, including computer-based and computer-adaptive tests (Brown 1997; Dunkel 1999; Yao and Ning 1998), audiotape-based oral proficiency interviews (Norris 1997; Stansfield and Kenyon 1992), and web-based testing (Roever 1998).

Recent discussion of so-called “alternative assessment” has highlighted the potential usefulness of a variety of innovative testing procedures, including portfolios, self- and peer-assessment, conferencing, diaries and learning logs, and teacher checklists and observations (Short 1993; TESOL Journal 1995). Increasing attention has also been given to performance-based and task-based language testing in which examinees are required to perform real-world tasks in the target language (Long and Norris in press; McNamara 1996; Norris, Brown, Hudson and Yoshioka 1998; Skehan 1998).

Faced with such an array of language testing alternatives, how are those people responsible for language testing to choose and use tests that are appropriate for their particular needs? Furthermore, on what basis can language teachers evaluate whether language tests are actually doing what they are supposed to be doing? In this article, I will offer three general recommendations to help language teachers and others select, use, and evaluate language tests and related alternatives in a purposeful manner. These recommendations are to (i) focus on assessment, not simply tests; (ii) clarify the intended use of the test; (iii) evaluate the outcomes of assessment.

Focus on assessment, not on tests

Language tests are simply instruments or procedures for gathering particular kinds of information, typically information having to do with students’ language abilities. Tests may have a variety of formats, lengths, item types, scoring criteria, and media. We may differentiate among language test types according to such characteristics and the information provided by each. For example, a 20-item cloze test, which asks the examinee to write single-word responses to complete a reading passage, provides a very different kind of information than does a 20-item multiple choice reading comprehension test, in which the examinee has only to choose the correct responses.

But deciding which of these test types is better or more appropriate is not easy. Knowing that each uses a unique format to provide different kinds of information does not bring us much closer to selecting one or the other alternative. Indeed, attempting to select the most appropriate among available testing alternatives on the basis of their characteristics alone would be like trying to choose between a hammer, a shovel, or a screwdriver based entirely on what these tools look like. Obviously, to select the appropriate tool, we first need to have an idea about the job to be accomplished. Alone, language tests tell us little about the jobs to be accomplished in language programs and classrooms. We cannot distinguish between good or bad, appropriate or inappropriate, reliable or not reliable, valid or not valid tests based solely on characteristics of the test instruments and procedures. Rather, we must focus instead on language assessment.

Language assessment is the process of using language tests to accomplish particular jobs in language classrooms and programs. In language assessment, we first gather information in a systematic way with the help of language testing tools. For example, we may use an oral interview to gather information about students' speaking abilities, then make interpretations based on that information. Or, we may make interpretations about students' abilities to perform a range of real-world speaking tasks based on how well students perform in the oral interview. Finally, based on these interpretations, we make a decision or take action within the classroom or program. We may decide that our students need more work on oral fluency and that we should therefore devote more class time to fluency-oriented activities.

Language assessment, then, is much more than simply giving a language test; it is the entire process of test use. Indeed, the ultimate goal of language assessment is to use tests to better inform us on the decisions we make and the actions we take in language education.

A number of such decisions and actions may call for the use of language tests. At the language program level, for example, we use tests to admit and place students into appropriate courses, determine the extent to which course objectives are being achieved, evaluate teacher performance, and reflect on the effectiveness of syllabus design and pedagogy. (For other uses of tests at the program level, see Brown 1995).

Within the language classroom we use tests to diagnose areas of learner need or sources of learning difficulties, reflect on the effectiveness of materials and activities, encourage student involvement in the learning process, track learner development in the L2, and provide students with feedback about their language learning progress for further classroom-based applications of language tests (Brown 1998; Cohen 1994).

We may also use language tests to influence or wash back on what and how we teach; make selection decisions for entrance into universities; and certify that students have achieved a particular standard of language ability.

Although not an exhaustive list, the examples make it clear that we use language tests to accomplish a wide range of jobs in language classrooms and programs. In order to use tests in a purposeful way, language teachers first need to establish exactly which jobs language assessment should be accomplishing within a particular language education context. With the jobs of

language assessment in focus, teachers will be better able to select the appropriate language testing tools and to decide how they should be used to get the jobs done.

Specify intended test use

Focusing on the aims of language tests is a crucial first step toward engaging in purposeful language assessment. However, simply knowing what decisions and actions will be based on language tests will probably not give us enough information to select the most appropriate testing tools. Indeed, the same assessment job may be accomplished in a number of different ways. Thus, in addition to knowing what job we want to accomplish with the help of a language test, we also need to give careful consideration to how the test will be used in our classrooms and programs.

In Figure 1 below, intended test use is defined as the interrelationship between four components of the language assessment process:

(a) who uses the test; (b) what information the test should provide; (c) why, or for what purpose, the test is being used; and (d) what consequences the test should have. For teachers to put together a clear idea of just how language tests should be functioning within their language classrooms and programs, each of the four questions in Figure 1 needs to be answered.

Figure 1 also shows that each of the four components interacts with all of the other components. Obviously, different people may want different information from tests for different purposes and with different consequences. In considering the interactions among these components of test use, teachers may realize just how differently language testing tools are being used within their particular language education contexts. However, by specifying exactly how tests are intended to be used within language classrooms and programs, language teachers should have a sound basis for evaluating and revising current assessment practice, for making the purposeful selection of appropriate language testing tools, for designating how those tools will be used, and for deciding to what extent language assessment is accomplishing the jobs it was intended to accomplish.

Who are the test users?

An important starting point for specifying intended test use is to clarify who uses the test-based information. In many language programs, teachers are the primary users of such information, because they are typically faced with making decisions and taking actions within the classroom on a daily basis. However, it is important not to overlook others who may use tests to make decisions or take actions. This list may include students, students' families, school administrators, curriculum planners, funding agencies, future employers, and university admissions officers.

Each of these potential users will naturally have particular reasons for looking at the information provided by tests, and they may require very different kinds of information from a test. Who uses test-based information will also determine how this information should be reported, as different test users may look for and understand different aspects of test score reports. Finally, different users will also attach varying degrees of importance to the outcomes of the language assessment process.

Obviously, then, the range of potential test users should be carefully considered before we select and implement test instruments or procedures. It will also be important to know who is using a test when we want to evaluate how well a test is accomplishing its intended job.

What is being tested?

In addition to knowing who uses test-based information, we need to identify exactly what information a test should provide. Although language tests may be used for a variety of reasons, and indeed, a single test may be used by different test users in very different ways, the amount and type of information provided by a language testing tool will always be limited. To ensure that the appropriate amount and type of information is gathered within the language assessment process, we first need to specify exactly what that information should be.

The most direct way to specify the information needed from language tests is to think about the kinds of interpretations test users will make. Naturally, the kinds of interpretations we make in language classrooms and programs are at least as numerous. Some common interpretations are students' global language proficiency, students' mastery of curricular objectives, students' grammatical knowledge, students' productive language abilities at the end of a course of study, the effectiveness of particular lessons, language learning objectives in need of attention within the classroom, and others. The exact interpretations will depend on the unique circumstances associated with particular language education contexts. One crucial aspect that we should note at this stage is that interpretations are always made about someone or something. We should therefore be careful to designate who or what we are interpreting.

Once we decide what interpretations we intend to make within the language assessment process, we can then specify exactly the amount and type of information needed to support these interpretations. For example, in order to support an interpretation about students' productive language abilities at the end of a course of study, we may conclude that we need information about both speaking and writing abilities. We may also decide that we will need to have information about students' abilities to perform several different speaking tasks and writing tasks that have been taught within our language curriculum.

With this specified amount and type of information, we should now be able to make a purposeful decision about the language testing tools we will need. We may conclude that multiple testing tools are able to provide us with the appropriate amount and type of information necessary to support the interpretations that we will make. Whatever conclusions we reach, the point at this stage is to make our selection of language testing tools more purposeful by specifying what interpretations test users will be making, deciding what information is necessary to support these interpretations, and deciding how many and what kinds of language tests will give us the required information.

What is the purpose of the test?

Just as we need to specify what information testing instruments and procedures should give us, we also need to make explicit exactly what purposes we have for using tests. Once we have administered language tests, we use the resulting information to inform various decisions and actions, that is, to accomplish the jobs of assessment.

It will prove helpful to give further consideration to just how we intend to use language assessment; that is, how we intend to use language tests to contribute to our overall classroom and program goals and our curricular objectives. What kinds of language testing tools will reflect the values underlying our program? Does our intended use of language tests contribute to accomplishing the curricular objectives in a manner that is consonant with our values and with the values of our students? How does language assessment in general fit into the big picture that we want our language program to portray? Answers to these questions should help us develop a better understanding of the overall purpose for using language tests within our language education contexts. Given explicit specification of exactly why we are engaging in assessment in the first place, we will be much better able to select and use language tests appropriate to our purposes.

What is the impact of the test?

As test users go about gathering information and making interpretations, which then lead to decisions and actions, the use of language tests obviously affects a variety of individuals and the language classrooms and programs themselves. We should therefore specify what consequences, both positive and negative, are likely to occur as a result of the intended use of language testing tools. Individuals who might be affected by using language tests include students, parents and other family members, teachers, employers, and others involved in the assessment process. Since each of these individuals may be influenced in different ways, we should consider the potential positive and negative consequences of a test. Additionally, the test may influence other elements of the language education context such as teaching practices, materials and activities, curricular objectives, enrollments, funding, program reputation, and so forth.

For all of the individuals and program elements that may be affected by our language assessment practice, we need to ask what the worst possible consequence of using the test is, what the ideal consequence of using the test is, and what the intended consequence of using the test is. Once we have specified the likely results of test use, we can select testing instruments and procedures that will maximize positive consequences and minimize negative consequences. Given an explicit statement about who or what the assessment practice should affect and in what ways, we can also have a basis for evaluating the actual consequences of test use.

Evaluate the outcomes of assessment

By specifying the intended uses for language tests, language teachers should be able to make well-motivated and purposeful selections of language testing instruments and procedures. However, regardless of how carefully we select and use language testing tools, we can never predict exactly what the outcomes of the assessment process will be. To keep our language assessment practice purposeful, we therefore need to evaluate the extent to which our language testing tools are actually helping to accomplish the jobs of language assessment in our classrooms and programs.

The in-depth evaluation of test use may involve a number of approaches and methods, including empirical investigations and observations (Messick 1989; Shepard 1997). Fundamentally, to evaluate how well our language testing tools are functioning, we need to reflect on the extent to which our use of language tests helps us to make the decisions and take the actions concerning

the jobs of language assessment. Thorough specifications of intended test use will provide a valuable resource.

In evaluating the outcomes of language test use, teachers should first reflect on the specifications of intended use, then decide the following: Were all of the actual test users identified? What range of interpretations did the test users actually make based on information provided by the language test? Did the testing tools provide the appropriate amount and type of information to support these interpretations? Were test-based decisions accurate and test-based actions appropriate? Were the intended classroom, curriculum, and program purposes fulfilled by using the language test? Finally, and perhaps most importantly, what were the actual positive and negative consequences of using the language test?

Based on the answers to these questions, we may decide that language testing tools are appropriate for our language assessment need. Or we may decide that test instruments and procedures need to be slightly revised, thoroughly overhauled, or thrown out.

In summary, purposeful language assessment involves the cyclical process of focusing on the jobs to be accomplished by language assessment, specifying the intended uses of language tests in accomplishing these jobs, selecting the appropriate language testing tools and designating how they are to be used, and evaluating the extent to which the jobs of language assessment are being accomplished with the help of language tests. Thus, to choose and use the language testing alternatives most appropriate for their language education contexts, language teachers need to keep in mind the purposeful nature of language assessment.

References

- Brown, J. 1995. *The elements of language curriculum: A systematic approach to program development*. Boston: Heinle and Heinle.
- . 1997. Computers in language testing: Present research and some future directions. *Language Learning and Technology*, 1, 1, pp. 44–59.
- . 1998. *New ways of classroom assessment*. Bloomington, IL: TESOL.
- Cohen, A. 1994. *Assessing language ability in the classroom*. 2nd ed. New York: Heinle and Heinle.
- Dunkel, P. 1999. Considerations in developing or using second/foreign language proficiency computer-adaptive tests. *Language Learning and Technology*, 2, 2, pp. 77–93.
- Long, M., and J. Norris. Forthcoming. Task-based language teaching and assessment. In *Encyclopedia of language teaching*. Ed. M. Byram. London: Routledge.
- McNamara, T. 1996. *Measuring second language performance: A new era in language testing*. New York: Longman.

Messick, S. 1989. Validity. In *Educational measurement*, 3rd ed. pp. 13–103. Ed. R. Linn. New York: American Council on Education.

Norris, J. 1997. The German speaking test: Utility and caveats. *Die Unterrichtspraxis*, 30, 2, pp. 148–158.

Norris, J., J. Brown, T. Hudson, and J. Yoshioka. 1998. *Designing second language performance assessments* (Technical Report 18). Honolulu, HI: University of Hawaii, Second Language Teaching and Curriculum Center.

Roever, C. 1998. *Web-based language testing* [HTML document].

<http://www2.hawaii.edu/~roever/wbt.htm>

Shepard, L. 1997. The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16, 2, pp. 5–13.

Short, D. 1993. Assessing integrated language and content instruction. *TESOL Quarterly*, 27, pp. 627–656.

Skehan, P. 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.

Stansfield, C., and D. Kenyon. 1992. Research on the comparability of the Oral Proficiency Interview and the Simulated Oral Proficiency Interview. *System*, 20, pp. 347–364.

TESOL Journal. 1995. Alternative assessment. Special issue. *TESOL Journal*, 5, 1.

Yao, T., and C. Ning. 1998. *Computer-adaptive test for reading Chinese* (CATRC) (Net Work 11) [HTML document]. Honolulu: University of Hawaii, Second Language Teaching and Curriculum Center.

Figure 1

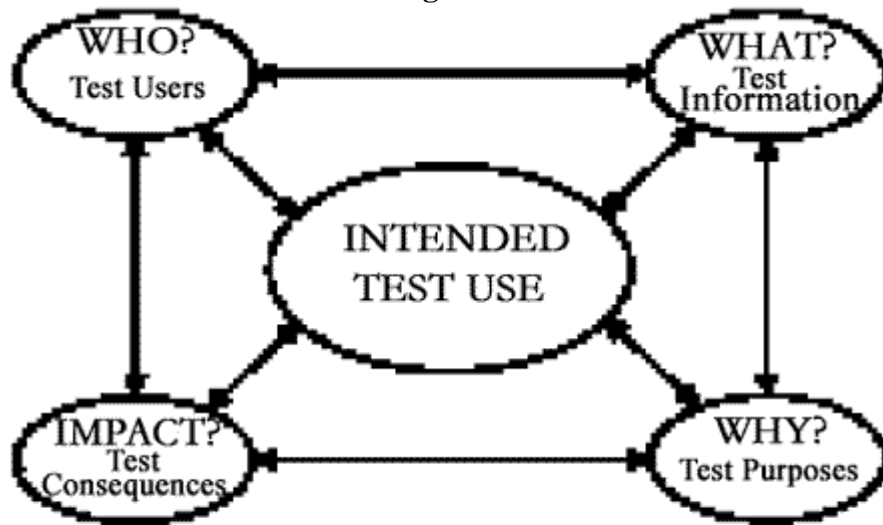


FIGURE 1

John Norris is a Mellon Fellow at the National Foreign Language Center Institute for Advanced Studies, Johns Hopkins University, USA.